Kinect V2 vs Intel RealSense D435: A Comparative Study on 3D Mapping

Lucas C. Nascimento* Waleson A. Melo* Emanuelle S. Gil* Alternei S. Brito* Felipe G. Oliveira*
José L. S. Pio†

*Institute of Exact Sciences and Technology (ICET), Universidade Federal do Amazonas (UFAM).
Itacoatiara, Amazonas, Brazil.
†Institute of Computing (ICOMP), Universidade Federal do Amazonas (UFAM).
Manaus, Amazonas, Brazil.
{luscarvalhox, walesonmelo23, manugil943}@gmail.com, {alternei, felipeoliveira}@ufam.edu.br,
josepio@icomp.ufam.edu.br

***Abstract*—The generation of accurate 3D maps is essential for an efficient autonomous navigation. With precise mapping, robots can plan optimal paths, avoiding obstacles and efficiently reaching their destinations. For this purpose, the use of RGB-D cameras, which capture color images and depth information of the environment, is common. This paper presents a comparative study of 3D maps generated by the Kinect V2 and RealSense D435 sensors, which were properly configured in a manipulator robot for controlled data acquisition, considering different environments and capture conditions. The RTAB-Map algorithm was used to process the data acquired by the sensors and generate the 3D maps. This analysis allows identifying which sensors are more suitable for each type of environment, as well as their limitations and advantages. Thus, this comparison helps select the best camera for each application and provides valuable insights for the development of more accurate and efficient applications.**

***Index Terms*—Three-dimensional mapping, 3D map comparison, RGB-D cameras, Comparative study**

I. Introduction

The evolution of technology has significantly contributed to the advancement of Robotics, allowing robots to be used in activities that require exceptional mobility skills. The use of mobile robots has been the subject of great effort in recent decades, both in industrial and academic scenarios, particularly regarding autonomous navigation. In order to perform more precise navigation tasks, the planning system of the robots needs to obtain accurate information about its surroundings, commonly provided by maps of the environment [1].

The mapping process, for mobile robots, involves the creation and representation of the robot's navigation environment. There are several different types of maps, such as metric, semantic, and topological maps. However, a more sophisticated mapping strategy consists of representing three-dimensional information around the scene through 3D maps. 3D maps enhance the robot's perception capabilities by providing depth and spatial information of the environment. In addition, it allows the robot to have a more accurate understanding of its surroundings, with detailed information about the shape, size and relative positions of objects and obstacles [2].

3D map generation is one of the most important technologies that allows robots to autonomously navigate indoors and outdoors. 3D maps are generated from sensors that capture information about the shape and geometry of objects and obstacles in an environment, as well as their color and texture features. Additionally, 3D maps can be updated in real-time as the robot moves through the environment. These maps allow robots to plan their routes accurately and avoid collisions with obstacles [3].

Different sensors can be used to generate 3D maps, such as LIDAR (Light Detection and Ranging) and SONAR (Sound Navigation and Ranging). These sensors use pulses of light and sound waves, respectively, for distance measurement and object detection. In addition to the mentioned sensors, RGB-D cameras are special devices that combine color images (RGB) with depth information (D). Depth can be achieved using techniques such as structured light and stereo vision, which involve projecting light patterns into the environment and measuring the time it takes for the light to return to the camera. The combination of visible and depth information allows for a more comprehensive representation of the environment compared to sensors that capture only color information [4].

This paper presents a comparative study of 3D maps generated by Kinect V2 and RealSense D435 sensors. For this purpose, the sensors were properly configured in a controlled environment for data acquisition, considering different real scenarios and data capture conditions. The RTAB-Map algorithm was used to process the data captured by the sensors and generate 3D maps. The same settings and parameters were considered to ensure a fair comparison. The 3D maps were compared based on the quality of the three-dimensional reconstruction, geometry fidelity, detail capture accuracy and map consistency.

II. Related work

The generation of 3D maps is of paramount importance for advancing technologies and applications that rely on accurate spatial understanding. 3D maps provide rich spatial

information, including geometry, texture and semantic details, enabling more advanced analysis and decision-making processes.

Many algorithms are used for 3D mapping, including the widely used RTAB-Map algorithm. In [5], the authors present a comparative study of the trajectories generated by the RealSense D435 and Kinect V2 sensors using the RTAB-Map algorithm. The evaluation was performed using the Root Mean Square Error (RMSE) of Euclidean distances between true and estimated trajectories. Through the experiments it was possible to observe that the RTAB-Map overcomes the Kintinuous technique, with RMSE of 0.0555. [6] compares three SLAM-based algorithms: RTAB-Map, ORB-SLAM2 and SPTAM. Simulations were performed indoors and outdoors, where RTAB-Map demonstrated more accurate estimates outdoors with stereo cameras, reaching an error rate of 4.54%. Finally, in [7], a comparative analysis of trajectories computed by different ROS-based SLAM systems was presented. The experiments were conducted in a typical office environment, and the results of the different methods were compared using appropriate metrics. Among the used algorithms, the RTAB-Map stands out, with the lowest RMSE value of 0.163.

Several researches have been developed for image and depth analysis using RGB-D cameras. Kinect and RealSense RGB-D cameras are widely used individually for mapping and detection problems, with promising results. In [8], the RGB-D Kinect camera was used to present a new approach to 3D reconstruction. The approach uses visual and geometric features, structure-from-motion techniques and innovative algorithms. Through experiments, including human head and body modeling, the effectiveness of the proposed approach was demonstrated in several challenging scenarios. In [9], the RGB-D RealSense D435 camera was used to capture infrared images. The authors proposed a new method for pedestrian detection and distance estimation using RGB-D data. Mask R-CNN was employed for segmentation and pedestrian detection, while the Semiglobal Matching technique was used to calculate depth maps from stereo images provided by the RealSense D435 camera. The results showed that the method achieved an average detection and distance estimation accuracy of 87.7%, within a range of 5 meters.

There are several works comparing RGB-D camera models. In [4], a study was carried out to compare the performance of 10 RGB-D sensors in 3D reconstruction. Using a controlled environment and a robotic arm, sensors were evaluated using RTAB-Map software. Astra Pro, Xtion, Astra, Kinect V2, and Kinect V1 sensors produced the best results, while older RealSense sensors underperformed due to range and technology limitations. The Kinect V2 had an error sum of 148.18cm and the RealSense D435 had 151.66cm.

[10] compares two RGB-D cameras, Intel RealSense D435 and Kinect V2, to measure the relative speed of a walking person. In this context, the Kinect V2 camera presents smaller standard deviations for the variables time (0.20), distance (0.014) and speed (0.10) compared to the Intel RealSense camera, indicating less dispersion and greater consistency in measurements. Thus, the authors suggest that the Kinect appears to be a more reliable RGB-D camera than the Intel RealSense D435 for collecting skeletal data when comparing capture range and signal quality.

In [11], a comparison was made between Kinect, RealSense and Xtion cameras using parameters such as RGB camera resolution, depth camera resolution, field of view, depth range, among others. According to the authors, the Kinect V2 was considered the best device in terms of data quality and applicability in larger robots.

III. Theoretical Background

The 3D map generation process plays a fundamental role in robotics field. Through detailed 3D maps of the environment, mobile robots gain a comprehensive understanding of their surroundings, allowing them to navigate and interact with the world more effectively. These maps provide crucial spatial information, including obstacle locations, object positions, and terrain features, increasing the robots capability of plan optimal paths and make better decisions in real-time.

Moreover, with the ability to create and update precise 3D maps during movement, mobile robots can navigate more efficiently, explore complex spaces and perform tasks with increased reliability and safety. In this work, is proposed a comparative study on 3D mapping, regarding Red, Green, Blue and Depth (RGB-D) cameras. For this, in this section, is presented the background about the used RGB-D cameras and the 3D mapping algorithm, concerning the mapping generation.

A. Kinect V2

The Kinect V2 is a device with an infrared depth sensor integrated with a high-resolution RGB camera, allowing for the simultaneous capture of color information and depth data (Figure 1). The Kinect was developed by Microsoft and uses the time-of-flight (ToF) technique to accurately measure the distance between the sensor and objects in the environment. The depth sensor of the Kinect V2 has a resolution of 512 x 424 pixels, and the RGB camera has a resolution of 1920 x 1080 pixels. As a result, it is possible to capture depth data and RGB images at a frame rate of up to 30 frames per second, providing real-time visualization.

Fig. 1: Kinect V2 sensor for capturing depth data and RGB images.

Due to its depth sensing capabilities, the Kinect V2 can generate detailed and accurate depth maps of the surrounding environment. These depth maps provide information about the geometric structure of the environment, enabling the creation

of 3D maps. By combining the depth data with RGB images, the Kinect V2 offers a comprehensive representation of the scene with real-time performance. As a result, the Kinect V2 has found applications in several fields, including robotics, virtual reality, and augmented reality. Its ability to generate detailed 3D maps facilitates navigation and localization in unknown environments, making it a valuable tool for applications that require spatial understanding and interaction in the 3D domain.

B. Real Sense D435

The Intel RealSense D435 is an advanced sensor for depth detection and visible imaging, designed for computer vision applications (Figure 2). With a stereo pair of RGB cameras and a structured light-based depth sensor, the RealSense D435 offers a combination of technologies that enable capturing high-quality images and accurate depth information. The sensor has a depth resolution of up to 1280 x 720 pixels, and the RGB camera has a resolution of up to 1920 x 1080 pixels. It can capture depth data and RGB images at frame rates of up to 90 frames per second, allowing for real-time visualization and fast performance. The RGB cameras provide sharp and detailed color images, while the depth sensor maps the three-dimensional geometry of the environment. Developed by Intel, it offers advanced capabilities for capturing data and generating detailed 3D maps of the surrounding environment.

Fig. 2: RealSense D435 with its pair of RGB stereo cameras and depth sensor.

The structured light technology of the RealSense D435 projects an infrared light pattern onto the environment and measures the deformation of this pattern to calculate the distance to objects. This allows the sensor to capture depth information with millimeter-level accuracy, providing a detailed representation of the environment. Its ability to capture real-time depth information and RGB images is extremely useful for robots to navigate safely and efficiently in unknown environments. The depth information provided by the sensor allows robots to determine the distance of objects in relation to themselves and it is used to generate 3D maps of the environment, which is essential for collision avoidance and safe path planning.

C. 3D Mapping Algorithm

The algorithms for 3D mapping are used to create three-dimensional representations of environments from data captured by sensors such as RGB-D cameras, LiDAR, or depth sensors. These algorithms process the information provided by the sensors and generate a 3D model of the environment, capturing the geometry, structure, and appearance of the objects in the scene. There are different approaches and techniques used in 3D mapping algorithms, depending on the sensors and specific application needs, such as voxelization, point cloud mapping, surface estimation, and SLAM (Simultaneous Localization and Mapping). These algorithms are essential for applications such as autonomous robot navigation, virtual reality, augmented reality, 3D environment reconstruction, among others. They allow systems to understand the structure of the environment, identify objects, plan trajectories, and interact with the three-dimensional world more accurately and efficiently. Among the techniques used to generate 3D maps from sensor data, we employ the RTAB-Map algorithm, which is particularly suitable for dynamic environments and has features that make it robust and efficient.

The RTAB-Map (Real-Time Appearance-Based Mapping) algorithm is a simultaneous localization and mapping (SLAM) algorithm that enables real-time creation of 3D maps while tracking the position and orientation of a robot or mobile sensor. Its robustness and ability to handle dynamic environments make it a popular choice in robotics applications and autonomous navigation systems. RTAB-Map employs an appearance-based approach for real-time 3D mapping, relying on visual feature matching to track the position and build the map. The basic operation of the RTAB-Map algorithm can be summarized in three main steps:

- Feature Extraction: RTAB-Map uses computer vision techniques to extract visual features from the image data captured by the sensor. This can include features such as interest points or relevant visual descriptors.
- Localization Tracking: The algorithm tracks the location of the robot or mobile sensor relative to an existing map by comparing the visual features extracted in the previous steps with the features stored in the map to determine the current position and orientation.
- Map Update: The algorithm updates the 3D map as new data is captured, incorporating depth information and other sensor data to build and expand the existing map. This may involve adding new points, updating texture information, or creating more detailed representations of the environment.

RTAB-Map also incorporates advanced techniques such as graph optimization and dynamic occlusion to improve mapping quality. It uses a graph to represent the structure of the environment and optimizes pose estimates and object placement based on available visual and depth information. An important feature of RTAB-Map is its ability to handle dynamic environments where objects may move or enter and exit the scene. It can identify and track moving objects, distinguishing between static and dynamic objects, thus avoiding the inclusion of mobile objects in the map.

IV. Experiments

A. Experimental setup

The experiments were carried out using: i) a RGB-D camera, Kinect V2; ii) a RGB-D camera, Intel RealSense D435; iii) an EPSON manipulator robot, with 6 degrees of freedom and precise movements; and iv) a Dell computer with an Intel Core^{TM} i7-8550U CPU and 32 GiB DDR3-2133 main memory. In this experimental setup, the Intel RealSense D435 is coupled above the Kinect V2 and both cameras are mounted on the EPSON manipulator robot. The EPSON robot is used in the experiments to provide precise and reproductible movements during the 3D map generation. The proposed experimental setup is illustrated in Figure 3.

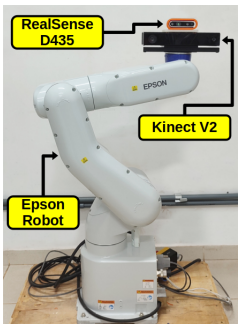
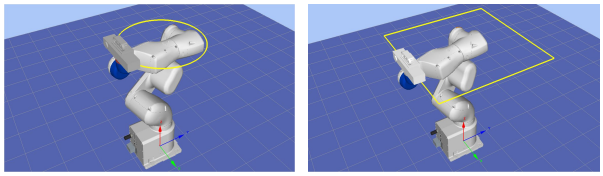


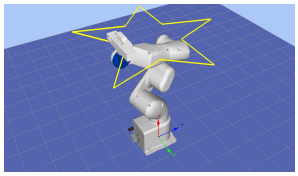
Fig. 3: Experimental setup used in the 3D map generating.

B. Experiment 1: Camera position estimation

Initially, it is important to highlight that, in all the experiments the 3D maps were generated by means of a set of controlled camera movements. We defined three controlled movements: i) circular shape; ii) square shape; and iii) star shape, as presented in Figure 4. Thereby, for each pre-defined controlled movement, were generated two 3D maps, one using Kinect V2 camera and other using Intel RealSense camera.



(a) (b)

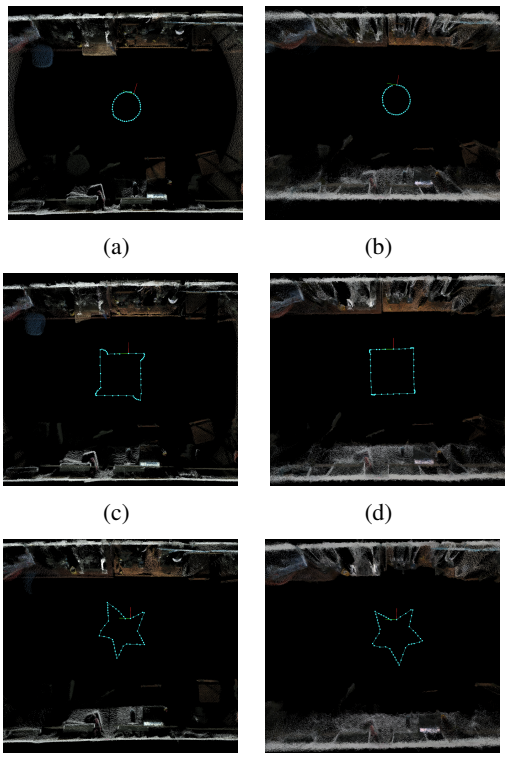


(c)

Fig. 4: Pre-defined camera movements performed by the Epson robot altogether the RGB-D cameras. Figures 4a, 4b and 4c represent controlled movements in circular, square and star shapes, respectively.

In the 3D mapping process an important step consists of to estimate the camera position, during the capture of the images and depths sequence. The estimated camera position is used to perform the multiple 3D maps merge, combining the individual 3D maps, in every camera position. Thus, an imprecise camera position estimation may lead to imprecise 3D maps.

In this sense, this experiment intend to compare the precision, in the camera position estimation task, during the 3D map generation, using the RTAB Map algorithm, through Kinect V2 and Intel RealSense cameras. For this, were used the three mentioned controlled movements (circle, square and star shapes). As can be seen in the Figure 5, from the known pre-defined movement positions and the estimated positions it is possible to calculate the precision of the estimated camera positions. For this, we used the Root Mean Square Error (RMSE) metric, where the smallest the values in Table I, the biggest the precision.



(a) (b)

(c) (d)

(e) (f)

Fig. 5: Examples of camera position estimation. Figures 5a and 5b correspond to the camera position estimation in circular movement, by Kinect V2 and RealSense, respectively. Figures 5c and 5d correspond to the camera position estimation in square movement, by Kinect V2 and RealSense, respectively. Figures 5e and 5f correspond to the camera position estimation in star movement, by Kinect V2 and RealSense, respectively.

From the achieved results presented above it is possible to verify, through qualitative and quantitative analysis, that the Intel RealSense camera is more precise than the Kinect V2

TABLE I: Precision in the camera position estimation using Kinect V2 and Intel RealSense cameras, regarding different controlled movements.

	Camera position evaluation	
Shapes	**Kinect V2**	**Intel RealSense**
Circle	0.6160	0.3076
Square	3.9183	1.5352
Star	4.3680	2.9633

camera, in the process of camera position estimation, for the 3D map generation.

C. Experiment 2: Density of points in the 3D map

For an effective and precise 3D map, the density of points is a paramount aspect. A 3D map is composed by a large amount of points, called point cloud. A dense point cloud provides a more detailed representation of the environment, capturing complex features and structures with higher accuracy. This level of detail is essential for applications such as autonomous navigation, object recognition, and scene understanding. Additionally, a dense point cloud improves the precision of measurements and calculations performed on the map. For tasks like distance estimation, surface reconstruction, or volumetric analysis, having more points densely distributed allows for more accurate and reliable results.

In this experiment we intend to evaluate the density of points in the generated 3D maps. For this, we computed the amount of points in the generated 3D maps by the RTAB-Map algorithm, using the Kinect V2 and Intel RealSense D435 cameras, in the different pre-defined controlled movements. In Table II are presented the amount of points for every generated 3D map.

TABLE II: Density of points in the generated 3D maps using Kinect V2 and Intel RealSense cameras, regarding different controlled movements.

	Amount of points in the 3D maps	
Shapes	**Kinect V2**	**Intel RealSense**
Circle	490363	768840
Square	540543	839019
Star	495256	827238

In this experiment, taking the obtained results in account, it is possible to verify that the Intel RealSense camera yields denser 3D maps, as we can observe in Table II, even with different camera movements in the acquisition process of images and depths.

D. Experiment 3: Geometric fidelity

Geometric fidelity is a fundamental quality aspect for mapping and reconstruction tasks. It ensures that the generated 3D model faithfully represents the current structure of the environment, allowing for accurate measurements, analysis, and visualization of the acquired data.

In this experiment we intend to evaluate the fidelity degree of the generated 3D maps, qualitatively and quantitatively. For this, a set of points was defined in the real-world environment and used to compare their positions and distances with the estimated 3D maps. Thereby, a quality metric can be used to calculate the accuracy rate between the known positions and distances and the estimated positions and distances, from the generated 3D maps. In Figure 6 we can see some examples of selected points in real-world, used in this experiment.

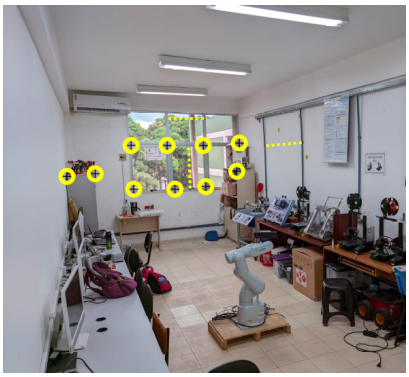


Fig. 6: Real-world environment used in the experiments, highlighting some selected points and distances.

In the qualitative analysis, regarding Figure 7, it is possible to observe some visual aspects related to the geometric fidelity of the estimated 3D maps. From the Figure 7 we can verify two main problems in the 3D maps generated using the intel realsense camera: i) curved surfaces, where the walls have a curved shape. It is possible to see that the further away from the camera, the less accurate the points are, in relation to the preservation of the flat structure of surfaces; and ii) inconsistent points, where the estimated points are plotted in wrong positions, farther than the expected position.

In the quantitative analysis, known selected points and distances are compared with real points and distances. For this, we used the RMSE quality metric, where the smallest the values in Table III, the biggest the precision. In Table III, it is possible to validate the high fidelity of the 3D maps generated by the Kinect V2 camera, using the RTAB-Map algorithm. From the qualitative and quantitative analysis we noticed that, although the intel realsense camera presented better results regarding camera position estimation and higher points density, the 3D maps produced by the Kinect V2 camera were more accurate and reliable, with greater geometric fidelity.

TABLE III: Geometric fidelity in the generated 3D maps using Kinect V2 and Intel RealSense cameras, regarding different controlled movements.

	Fidelity degree in the 3D maps	
Shapes	**Kinect V2**	**Intel RealSense**
Circle	1.2627	3.3068
Square	0.5783	2.4514
Star	0.0949	2.3119

V. Conclusion and Future Work

In this paper, a comparative study of 3D maps generated by the Kinect V2 and Intel RealSense D435 sensors using

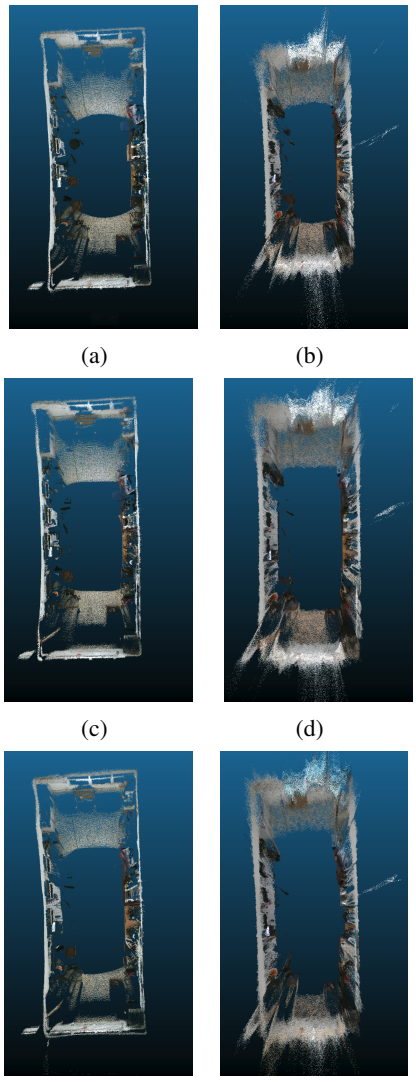


Fig. 7: 3D maps generated using Kinect V2 and RealSense cameras, regarding different movements, by the RTAB-Map algorithm. Figures 7a and 7b correspond to 3D maps generated in circular movement, by Kinect V2 and RealSense, respectively. Figures 7c and 7d correspond to 3D maps generated in square movement, by Kinect V2 and RealSense, respectively. Figures 7e and 7f correspond to 3D maps generated in star movement, by Kinect V2 and RealSense, respectively.

the RTAB-Map algorithm was presented. Real-world experiments were conducted in a controlled environment to acquire data, considering different real scenarios and data capture conditions. The comparison was based on the quality of the 3D reconstruction, regarding the fidelity of the geometry and accuracy in capturing details.

The Kinect V2 camera outperformed the Intel RealSense D435 in terms of the point cloud quality, presenting a consistent point cloud. Regarding performance in different environments, it achieved better results in all tested environments. However, it showed less flexibility as it relied on an external power source, making it more suitable for applications that require higher precision and have fixed locations.

The RealSense D435 camera presented noisy point clouds which calculates points in inconsistent positions, like beyond walls. However, estimates accurate camera positions during the 3D map generation. Additionally, due to its smaller size and independence from an external power source, the RealSense D435 is more suitable for mobile devices such as robots and drones. Its adaptability and better integration in mobile environments are important advantages in these scenarios.

In future work, the comparative analysis will be expanded to include other types of RGB-D cameras, evaluating the performance of different sensors. Furthermore, the combination of maps generated by these sensors will be explored to achieve greater accuracy. Additionally, the sensors will be used altogether with a manipulator robot for object grasping purposes and mobile robots for autonomous navigation.

ACKNOWLEDGMENTS

This work was developed with support from CAPES/Brazil - Finance Code 001, CNPq/Brazil - Grant 409109/2021-5 and Motorola, through the IMPACT-Lab R&D project, in the Institute of Computing (ICOMP) of the Federal University of Amazonas (UFAM)

REFERENCES

[1] S. M. Mehdi, R. A. Naqvi, and S. Z. Mehdi, "Autonomous object detection and tracking robot using kinect v2," in *2021 International Conference on Innovative Computing (ICIC)*, 2021, pp. 1–6.

[2] G. Brahmanage and H. Leung, "Outdoor rgb-d mapping using intel-realsense," in *2019 IEEE SENSORS*, 2019, pp. 1–4.

[3] R. Fernandes, T. L. Rocha, H. Azpúrua, G. Pessin, A. A. Neto, and G. Freitas, "Investigation of visual reconstruction techniques using mobile robots in confined environments," in *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*, 2020, pp. 1–6.

[4] J. G. da Silva Neto, P. J. da Lima Silva, F. Figueredo, J. M. X. N. Teixeira, and V. Teichrieb, "Comparison of rgb-d sensors for 3d reconstruction," in *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, 2020, pp. 252–261.

[5] N. Altuntaş, E. Uslu, F. Çakmak, M. F. Amasyalı, and S. Yavuz, "Comparison of 3-dimensional slam systems: Rtab-map vs. kintinuous," in *2017 Int. Conf. on Comp. Science and Eng.*, 2017, pp. 99–103.

[6] K. J. de Jesus, H. J. Kobs, A. R. Cukla, M. A. de Souza Leite Cuadros, and D. F. T. Gamarra, "Comparison of visual slam algorithms orb-slam2, rtab-map and sptam in internal and external environments with ros," in *2021 Latin American Robotics Symposium (LARS), 2021 Brazilian Symposium on Robotics (SBR), and 2021 Workshop on Robotics in Education (WRE)*, 2021, pp. 216–221.

[7] M. Filipenko and I. Afanasyev, "Comparison of various slam systems for mobile robot in an indoor environment," in *2018 International Conference on Intelligent Systems (IS)*, 2018, pp. 400–407.

[8] K. Wang, G. Zhang, and H. Bao, "Robust 3d reconstruction with an rgb-d camera," *IEEE Trans. on Image Processing*, 2014.

[9] A. Tupper and R. Green, "Pedestrian proximity detection using rgb-d data," in *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2019, pp. 1–6.

[10] J. D. Mejia-Trujillo, Y. J. Castaño-Pino, A. Navarro, J. D. Arango-Paredes, D. Rincón, J. Valderrama, B. Muñoz, and J. L. Orozco, "Kinect™ and intel realsense™ d435 comparison: a preliminary study for motion analysis," in *2019 IEEE Int. Conf. on E-health Networking, Application Services (HealthCom)*, 2019.

[11] R. Zou, X. Ge, and G. Wang, "Applications of rgb-d data for 3d reconstruction in the indoor environment," in *2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*, 2016, pp. 375–378.